A GENERALIZED APPROACH TO POSSIBILISTIC CLUSTERING ALGORITHMS

JIAN ZHOU
Department of Industrial Engineering, Tsinghua University,
Beijing 100084, China
jian@tsinghua.edu.cn

CHIH-CHENG HUNG
School of Computing and Software Engineering,
Southern Polytechnic State University, Marietta, Georgia 30060, USA
chung@spsu.edu

Received 10 September 2006
Revised 2 November 2006
Accepted 30 November 2006

Fuzzy clustering is an approach using the fuzzy set theory as a tool for data grouping, which has advantages over traditional clustering in many applications. Many fuzzy clustering algorithms have been developed in the literature including fuzzy c-means and possibilistic clustering algorithms, which are all objective-function based methods. Different from the existing fuzzy clustering approaches, in this paper, a general approach of fuzzy clustering is initiated from a new point of view, in which the memberships are estimated directly according to the data information using the fuzzy set theory, and the cluster centers are updated via a performance index. This new method is then used to develop a generalized approach of possibilistic clustering to obtain an infinite family of generalized possibilistic clustering algorithms. We also point out that the existing possibilistic clustering algorithms are members of this family. Following that, some specific possibilistic clustering algorithms in the new family are demonstrated by real data experiments, and the results show that these new proposed algorithms are efficient for clustering and easy for computer implementation.

Keywords: Fuzzy clustering; possibilistic clustering; fuzzy set theory.

1. Introduction

Cluster analysis is the process of partitioning a data set into subsets of objects which have similar properties. Clustering methods have been used extensively in computer vision and pattern recognition. Fuzzy clustering is an approach using the fuzzy set theory as a tool for data grouping, which has advantages over traditional clustering in many applications since a flexible partition would be provided as a result by fuzzy clustering which is usually more proper in many problems.
In the literature on fuzzy clustering, the fuzzy c-means (FCM) clustering algorithm developed by Bezdek is the most well-known and used method.\textsuperscript{1} The FCM algorithm and its derivatives have been successfully used in many applications. However, the memberships resulting from FCM do not always correspond to the intuitive concept of degrees of belongingness or compatibility due to the probabilistic constraints that the memberships of each data point across classes must sum to 1. As a result, it may be inaccurate in a noisy environment.\textsuperscript{2}

To improve this weakness of FCM, Krishnapuram and Keller relaxed the probabilistic constraints and proposed possibilistic clustering algorithms (PCAs),\textsuperscript{2,3} where the memberships provide a good explanation of degrees of belongingness for the data. Since the clustering performance of the PCAs heavily depends on the parameters used,\textsuperscript{2,3} Yang and Wu suggested a new possibilistic clustering algorithm whose performance can be easily controlled.\textsuperscript{4} Compared with FCM, the PCAs are more robust to noise and outliers. Thus the PCAs have been applied to problems such as shell clustering, boundary detection, surface and function approximations.\textsuperscript{5–7}

Both FCM and PCAs are iterative algorithms, in which the update equations for memberships and cluster centers are all derived from the necessary conditions for a minimizer of some objective functions. Since each cluster is assumed to be a fuzzy set in fuzzy clustering, it is natural to provide an appropriate membership function for each cluster according to some given criteria from the fuzzy set theory and then evaluate the memberships of points belonging to clusters directly using the membership functions obtained.

Based on this idea, our paper presents a new approach of fuzzy clustering to generalize the existing possibilistic clustering algorithms, in which the update of cluster centers is performed via a performance index after the calculation of memberships using the fuzzy set theory. This generalized method leads to an infinite family of possibilistic clustering algorithms. We also point out that the existing PCAs are the members of this new family. Following that, some real data clustering performance on the basis of some specific algorithms in the family together with the three existing PCAs are proposed to demonstrate their efficiency. Furthermore, a numerical comparison with FCM on the accuracy and real cluster number of clustering results reveals the weakness of the PCAs.

This paper is organized as follows. In Section 2, the fundamental of fuzzy clustering is introduced using the fuzzy set theory and a brief review on fuzzy clustering is given involving FCM and three PCAs. Then, a generalized approach of possibilistic clustering is developed in Section 3, which leads to an infinite family of PCAs. Section 4 presents some computational experiments based on some data sets with some specific algorithms in the family as well as the three existing PCAs and FCM, whose results show the advantage and disadvantage of the PCAs. Finally, some conclusions and future research on the possibilistic clustering algorithms are given in Section 5.
2. Fuzzy Clustering

Since its introduction in 1965 by Zadeh, fuzzy set theory has been well developed and applied in a wide variety of real problems. In clustering, a great deal of research on fuzzy clustering has been accomplished in the literature. In this section, the fundamental of fuzzy clustering is introduced and fuzzy clustering algorithms including FCM and three PCAs are briefly reviewed.

2.1. Fundamental of fuzzy clustering

As a prerequisite for further discussion, notations used in this paper are given in advance:

- \( X = \{ x_j \mid j \in J \} \): a data set in a \( p \)-dimensional Euclidean space \( \mathbb{R}^p \) with its ordinary Euclidean norm \( \| \cdot \| \), where \( J = \{1, 2, \ldots, n\} \) is the index set of data;
- \( c \): a positive integer with \( c > 1 \) representing the number of clusters;
- \( \Theta = \{ \Theta_1, \Theta_2, \ldots, \Theta_c \} \): set of clusters \( \Theta_i \) representing the \( i \)th cluster, \( i \in I = \{1, 2, \ldots, c\} \);
- \( A = (a_1, a_2, \ldots, a_c) \): cluster center matrix, where \( a_i \) (vector) represents the \( i \)th cluster center for each \( i \in I \).

Fuzzy clustering is an approach using the fuzzy set theory as a tool for data grouping. As a basic assumption in fuzzy clustering, each cluster \( \Theta_i \) is supposed to be a fuzzy set with membership function \( \mu_i \) in each iteration, where \( \mu_i \) is defined on the set \( X \) by

\[
\mu_i(x_j) = \mu_{ij} \quad \text{for each } x_j \in X, \quad (1)
\]

with the variable \( \mu_{ij} \) representing degree of compatibility or membership of feature point \( x_j \) belonging to fuzzy cluster \( \Theta_i \). According to the fuzzy set theory, the memberships \( \mu_{ij} \) assigned should satisfy

\[
0 \leq \mu_{ij} \leq 1 \quad \text{for all } (i, j) \in (I, J). \quad (2)
\]

For convenience, denote the membership matrix by \( \mu \) with

\[
\mu = \begin{pmatrix}
\mu_{11} & \cdots & \mu_{1n} \\
\vdots & \ddots & \vdots \\
\mu_{c1} & \cdots & \mu_{cn}
\end{pmatrix}, \quad (3)
\]

and denote the set of all the membership matrices satisfying (2) by \( U_X \) with

\[
U_X = \{ \mu \mid 0 \leq \mu_{ij} \leq 1 \text{ for all } (i, j) \in (I, J) \}, \quad (4)
\]

called the eligible membership matrix set of \( X \).

The objective in fuzzy clustering is to find the most possible membership matrix \( \mu \in U_X \) according to the given data information and some decision criteria from experts’ opinions.
2.2. Fuzzy c-means algorithm

In fuzzy clustering, the FCM algorithm developed by Bezdek is the best known and used method, which is an iterative algorithm. After random initializations of cluster centers, the membership matrix $\mu$ and cluster center matrix $A$ are updated until the convergence of cluster centers, where the update equations for both $\mu$ and $A$ are obtained from the necessary conditions for a minimizer of objective function $J_{FCM}$ with

$$ J_{FCM}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \|x_j - a_i\|^2 $$

subject to the constraints $\mu \in U_X$ and

$$ \sum_{i=1}^{c} \mu_{ij} = 1 \quad \text{for all } j \in J. $$

The weighting exponent $m > 1$ is called the fuzzifier which has a great influence on the performance of fuzzy clustering. The update equations for $\mu$ and $A$ are

$$ \mu_{ij} = \left( \sum_{k=1}^{c} \frac{\|x_j - a_k\|^{2/(m-1)}}{\|x_j - a_k\|^{2/(m-1)}} \right)^{-1} \quad \text{for } (i,j) \in (I,J) $$

and

$$ a_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m x_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \quad \text{for } i \in I, $$

respectively.

The probabilistic constraints (6) are not appropriate to restrict the memberships according to the fuzzy set theory, consequently the memberships resulting from FCM do not always correspond to the intuitive concept of degree of belongingness or compatibility. In order to correct this weakness, possibilistic clustering algorithms were developed by relaxing the probabilistic constraints and designing other objective functions.

2.3. Possibilistic clustering algorithms

In order to produce memberships with a good explanation of the degrees of belongingness for the data, the PCAs were proposed by Krishnapuram and Keller. After relaxing the probabilistic constraints (6) in FCM, two objective functions $J_{PCA93}$ and $J_{PCA96}$ were created for the PCAs (PCA93 and PCA96), respectively, with

$$ J_{PCA93}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m \|x_j - a_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1 - \mu_{ij})^m $$

and

$$ J_{PCA96}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij} \|x_j - a_i\|^2 + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (\mu_{ij} \ln(\mu_{ij} - \mu_{ij})). $$
The necessary conditions for a minimizer \((\mu, A)\) of \(J_{PCA93}\) and \(J_{PCA96}\) subject to the constraint \(\mu \in U_X\) are the following update equations for memberships in PCA93 and PCA96, respectively,

\[
\begin{align*}
\mu_{ij} &= \frac{1}{1 + \left(\frac{||x_j - a_i||^2}{\eta_i}\right)^{1/(m-1)}} \quad \text{for} \ (i, j) \in (I, J) \quad \text{(PCA93)} \\
\mu_{ij} &= \exp\left\{-\frac{||x_j - a_i||^2}{\eta_i}\right\} \quad \text{for} \ (i, j) \in (I, J) \quad \text{(PCA96)}
\end{align*}
\]

and the update equation (8) for cluster centers in both PCA93 and PCA96, which is the same as those in FCM. The parameters \(\eta_i\) were recommended by Krishnapuram and Keller, as

\[
\eta_i = K \frac{\sum_{j=1}^{n} (\mu_{ij})^m ||x_j - a_i||^2}{\sum_{j=1}^{m} (\mu_{ij})^m} \quad \text{or} \quad \eta_i = \frac{\sum_{i,j} ||x_j - a_i||^2}{\sum_{i,j} 1}
\]

for \(i \in I\), where the parameter \(K > 0\) is typically chosen to be one, and \(\alpha\) is a parameter with \(0 < \alpha < 1\).

Since the possibilistic clustering performance heavily depends on the chosen parameters \(\eta_i\), a new PCA was developed by Yang and Wu (PCA06) whose performance can be easily controlled. The objective function used in PCA06 is

\[
J_{PCA06}(\mu, A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m ||x_j - a_i||^2 + \frac{\beta}{m^2 \sqrt{c}} \sum_{i=1}^{c} \sum_{j=1}^{n} [(\mu_{ij})^m \ln(\mu_{ij})^m - (\mu_{ij})^m],
\]

and accordingly the update equation for memberships is

\[
\mu_{ij} = \exp\left\{-\frac{m \beta ||x_j - a_i||^2}{\beta}ight\} \quad \text{for} \ (i, j) \in (I, J),
\]

where the parameter \(\beta\) is a measure of separation degree of the data set, defined by

\[
\beta = \frac{\sum_{j=1}^{n} ||x_j - \bar{x}||^2}{n} \quad \text{with} \ \bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}.
\]

The same update equation (8) for cluster centers in FCM is also used in PCA06. Compared with PCA93 and PCA96, PCA06 could provide good clustering results with higher accuracy through simple performance in most situations, which can be seen from the numerical experiments in Section 4.

3. Possibilistic Clustering: A Generalized Approach

Both FCM and the PCAs are iterative algorithms based on some objective functions. The existing possibilistic clustering algorithms provided some methods to
compute the membership of each point belonging to clusters using some special objective functions. However it is clear that many more functions could be the membership function of each cluster besides the functions suggested in the existing PCAs. In this section, a new method of fuzzy clustering is suggested to generalize the existing PCAs by calculating the memberships using the fuzzy set theory, and updating the cluster centers via a performance index. A new infinite family of PCAs, which involves the existing PCAs as members, is then obtained from the derivation.

### 3.1. Membership function in fuzzy set theory

In fuzzy clustering, the variable $\mu_{ij}$ is used to denote the membership degree of feature point $x_j$ belonging to fuzzy cluster $\Theta_i$ for each $(i, j) \in (I, J)$. Instead of deriving the update equations for $\mu_{ij}$ from some objective functions like the existing PCAs, a good idea is to evaluate the memberships directly according to the data information using the fuzzy set theory in each iteration. It is natural to provide an appropriate evaluation criterion for memberships, the fuzzy set theory is reviewed for this purpose.

By employing the membership functions defined on a universal set of discourse, the fuzzy set theory can represent ill-defined classes and concepts in a natural way. For any fuzzy set $\tilde{A}$, there is a standard or ideal of $\tilde{A}$, which is a semantic definition of the central concepts. Denote the standard of $\tilde{A}$ as $x_0$. The valuation of the membership $\mu(x)$ could be regarded as the comparison of element $x$ in the universal set of discourse with the standard $x_0$. Suppose that the comparison results in a dissimilarity $d(x, x_0)$ between $x$ and $x_0$. Then a monotone decreasing function of the dissimilarity $d(x, x_0)$ could be a good measurement for membership $\mu$, which implies the smaller the dissimilarity $d(x, x_0)$ is, the larger the membership value $\mu(x)$, i.e., the higher the grade of membership of $x$ in $\tilde{A}$, which is an intuitive result. Moreover, if the element $x$ has all the typicality of the standard $x_0$, the dissimilarity $d(x, x_0)$ should be zero, and accordingly the membership of $x$ belonging to fuzzy set $\tilde{A}$ is unity. If there is no similarity between $x$ and $x_0$, the dissimilarity should be infinite, and the corresponding membership value is zero.

The theory above can be applied in fuzzy clustering to compute the memberships in each iteration. It is natural to select the cluster center $a_i$ as the standard of fuzzy cluster $\Theta_i$ and measure the dissimilarity between point $x_j$ and the standard $a_i$ by the distance $||x_j - a_i||$. Thus for each fuzzy cluster $\Theta_i$, a monotone decreasing function $f_i$ of the distance could be a good substitution for the membership function $\mu_i$ defined in (1). In addition, it follows from above that $f_i(0) = 1$ and $f_i(+\infty) = 0$.

In summary, in each iteration of fuzzy clustering, the memberships $\mu_{ij}$ would be evaluated by

$$
\mu_{ij} = f_i(||x_j - a_i||) \quad \text{for } (i, j) \in (I, J),
$$

(16)
where the membership function $f_i$ satisfies the constraints
\[
\begin{align*}
    f_i &\text{ is monotone decreasing on } [0, +\infty) \\
    f_i(0) &= 1 \\
    f_i(+\infty) &= 0
\end{align*}
\] (17)
for each $i \in I$. It is easy to verify that the membership matrix obtained by (16) and (17) is in the eligible membership matrix set $U_X$ due to $||x_j - a_i|| \geq 0$. If $f_i$ satisfy the constraints (17) for all $i \in I$, we say the function vector $f = (f_1, \ldots, f_c)$ satisfies (17). Especially, if all the fuzzy clusters in $\Theta$ are known to have the similar membership properties, an identical function $f$ may be used for all the clusters, and then the memberships $\mu_{ij}$ are simply calculated by
\[
    \mu_{ij} = f(||x_j - a_i||) \quad \text{for } (i, j) \in (I, J).
\] (18)

In this case, we simplify the function vector $(f, f, \ldots, f)$ as $f$. It is obvious that there are many functions satisfying (17), for example,
\[
    g(x) = \frac{1}{1 + kx^b} \quad \text{or} \quad a^{-kx^b}, \quad a > 1, b > 0, k > 0,
\] (19)
some of which are illustrated in Figure 1. We would also like to point out that the update equations (11) and (14) for memberships in PCA93, PCA96 and PCA06 also satisfy (16) and (17) under the assumption that the parameters $\eta_i$ are positive constants in each iteration.\(^2\) It is easy to rewrite the membership function vectors of PCA93, PCA96 and PCA06 as $(f_1^{\text{PCA93}}, \ldots, f_c^{\text{PCA93}})$, $(f_1^{\text{PCA96}}, \ldots, f_c^{\text{PCA96}})$ and $f^{\text{PCA06}}$, respectively, where
\[
    f_i^{\text{PCA93}}(x) = \frac{1}{1 + (x^2/\eta_i)^{1/(m-1)}}, \quad i \in I,
\] (20)
\[
    f_i^{\text{PCA96}}(x) = \exp \left\{ -\frac{x^2}{\eta_i} \right\}, \quad i \in I
\]
and
\[
    f^{\text{PCA06}}(x) = \exp \left\{ -\frac{m\sqrt{\beta}}{\beta} x^2 \right\}.
\] (21)

Generally speaking, the evaluation equation (16) for memberships could be different for all the iterations. Thus, a more general expression of evaluation equations for memberships is
\[
    \mu_{ij}^{(t)} = f_i^{(t)}(||x_j - a_i||) \quad \text{for } (i, j) \in (I, J),
\] (22)
where $t$ is the iteration counter, and the functions $f_i^{(t)}$ satisfy the constraints (17). In PCA93\(^2\) and PCA96,\(^3\) it was recommended to reestimate the parameters $\eta_i$ in each iteration if the actual shape of the possibility distributions generated is important. In other words, different membership functions $f_i^{(t)\text{PCA93}}$ and $f_i^{(t)\text{PCA96}}$ are used for the $t$th iteration in the situations.
3.2. Cluster center update equation

In order to obtain the update equation for cluster centers, the performance-index approach is employed. It is a natural idea to use a sum of weighted squared error index $J_{FC}$ as a measure of the compactness of data with

$$J_{FC}(A) = \sum_{i=1}^{c} \sum_{j=1}^{n} w_{ij} ||x_j - a_i||^2. \quad (23)$$

When setting the $m$th power of membership $\mu_{ij}$ as the weight of distance $||x_j - a_i||$ in the performance index $J_{FC}$, a new index $J_{GPC}$ is obtained with

$$J_{GPC}(A) = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij})^m ||x_j - a_i||^2, \quad (24)$$

where $m > 1$ is a fuzzifier as defined in FCM. The update equation for cluster centers is derived from the necessary conditions for solutions of minimizing $J_{GPC}$, i.e.,

$$a_i = \frac{\sum_{j=1}^{n} (\mu_{ij})^m x_j}{\sum_{j=1}^{n} (\mu_{ij})^m} \quad \text{for } i \in I \quad (25)$$

which is the same as (8) in FCM.

**Remark 1.** It should be noted that except the $m$th power of $\mu_{ij}$ used in $J_{GPC}$, other appropriate weights can also be used in $J_{FC}$ to result in various update equations for cluster centers. For example, if we set the $m$th power of $C_{rij}$ as the weights of the distance $||x_j - a_i||^2$, a new approach is obtained, called credibilistic
clustering approach, where $\mathcal{C}_{ij}$ represents the credibility of the fuzzy event that the point $x_j$ belongs to the cluster $a_i$, $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, m$. The credibility measure plays the role of probability measure, and credibility theory is a branch of mathematics that studies the behavior of fuzzy event. The interested reader may consult the books by Liu for detailed expositions.\textsuperscript{13,14}

### 3.3. Generalized possibilistic clustering algorithms

By computing the memberships by the equation (22) and then updating the cluster centers according to the equation (25), a generalized approach to possibilistic clustering is developed, which is called generalized possibilistic clustering in the paper. The algorithms of generalized possibilistic clustering are summarized as follows.

#### Generalized Possibilistic Clustering Algorithms

**Step 0** Initialize $a_i^{(0)} \in \mathbb{R}^p$ for all $i \in I$, and set a small number $\epsilon > 0$ and iteration counter $t = 0$.

**Step 1** Compute $\mu_i^{(t+1)}$ using the evaluation equation (22) for all $(i, j) \in (I, J)$, where the membership functions $f_i^{(t+1)}$ satisfying the constraints (17) are predetermined.

**Step 2** Compute $a_i^{(t+1)}$ using the update equation (25) for all $i \in I$.

**Step 3** Increment $t$ until $\max_i ||a_i^{(t+1)} - a_i^{(t)}|| < \epsilon$.

Note that the only difference among all the PCAs derived from generalized possibilistic clustering is the function vector $f^{(t)} = (f_1^{(t)}, \ldots, f_c^{(t)})$ used in Step 1. A series of function vectors $\{f^{(t)}\}$ satisfying (17) will establish a corresponding PCA, denoted as $\text{PCA}\{f^{(t)}\}$. Consequently, a new infinite family $F_{\text{PCA}}$ of PCAs is obtained from the derivation with

$$F_{\text{PCA}} = \{\text{PCA}\{f^{(t)}\} \mid f^{(t)} \text{ satisfies the constraints (17)}\}. \quad (26)$$

When an identical function vector $f$ is used for all the iterations, the relevant PCA is written as $\text{PCA}\{f\}$. It is obvious that the existing PCAs involving $\text{PCA93}$, $\text{PCA96}$ and $\text{PCA06}$ are all members of the new family $F_{\text{PCA}}$. They are indeed $\text{PCA}\{f_1^{\text{PCA93}}, \ldots, f_c^{\text{PCA93}}{)}^{(t)}\}$, $\text{PCA}\{f_1^{\text{PCA96}}, \ldots, f_c^{\text{PCA96}}{)}^{(t)}\}$ and $\text{PCA}\{f^{\text{PCA06}}\}$, respectively.

**Remark 2.** Since it is not easy to get the information on the membership qualities of clusters in most real applications, it is recommended to use the same function for all the clusters in the real clustering performance by the PCAs. Though valuating the memberships $\mu_{ij}$ by appropriate functions for various iterations would provide better clustering results sometimes, however, it is not easy to decide the appropriate membership functions for each iteration by analyzing the data sets. Thus the possibilistic clustering algorithms $\text{PCA}\{f\}$ with an identical membership function
for all the clusters in all the iterations should be a good choice for clustering, which make the clustering process simpler to accomplish and easier to control.

In the following section, some real data experiments based on the possibilistic clustering algorithms like PCA\{f\} will be presented to illustrate the performance and efficiency of this specific type of PCAs in the family \( F_{PCA} \) compared with the existing three PCAs and FCM.

4. Computational Experiments

In this section, three real data sets involving the Iris data set,\(^{15,16}\) the glass and vowel data sets,\(^{17}\) are used for clustering. In these data sets, the real labels of points are known. We implement the clustering performance by processing PCA93, PCA96, PCA06, some new possibilistic clustering algorithms like PCA\{f\} in the family \( F_{PCA} \), and FCM. In order to see the influences of fuzzifier on all the clustering algorithms, the parameter \( m \) is taken as 2.0, 2.5 and 3.0, respectively. In order to compare the results obtained by running these clustering algorithms, some clustering mapping and evaluation criteria for clustering results are presented first for this purpose.

4.1. Clustering mapping criterion

Suppose that a real data set \( X \) is given, in which a label \( l_j \in I \) is assigned to each point \( x_j \) in \( X \) for \( j \in J \), which means that the point \( x_j \) belongs to the cluster \( \Theta_{l_j} \) in the real situation. After any clustering algorithm is run for the data set \( X \), some cluster centers \( \mathbf{a}_1^*, \ldots, \mathbf{a}_c^* \) corresponding to the clusters \( \Theta_1^*, \ldots, \Theta_c^* \) are provided. By distributing each point to the nearest cluster center in \( \{\mathbf{a}_1^*, \ldots, \mathbf{a}_c^*\} \), each point \( x_j \) in \( X \) is labelled with a new index \( r_j \in I \) for \( j \in J \), and a clustering result \( \mathbf{r} = (r_1, \ldots, r_n) \) is then obtained.

Generally the cluster \( \Theta_i^* \) in a clustering result does not correspond to the cluster \( \Theta_i \) in the real data set \( X \). Thus before valuating a clustering result \( \mathbf{r} \), we must decide the mapping between \( \{\Theta_1, \ldots, \Theta_c\} \) and \( \{\Theta_1^*, \ldots, \Theta_c^*\} \). In the real clustering performance, a “maximum mapping number criterion” is always adopted by most researchers. Let us illustrate the criterion through a simple example as follows.

Example 1. Suppose that a clustering result \( \mathbf{r}^* \) of a data set \( X \) with \( n = 100 \) points and \( c = 3 \) clusters is given. Count the number of points labelled with \( i \) in the real data \( X \) and assigned with \( k \) in the result \( \mathbf{r}^* \) for each \( i, k \in \{1, 2, 3\} \), and denote the number as \( s_{ik} \). Then a \( 3 \times 3 \) matrix \( S \) is obtained with

\[
S = \begin{pmatrix}
s_{11} & s_{12} & s_{13} \\
 s_{21} & s_{22} & s_{23} \\
 s_{31} & s_{32} & s_{33}
\end{pmatrix}.
\]  \hspace{1cm} (27)
The numbers $s_{ik}$ are called mapping numbers and $S$ is called the mapping matrix of the clustering result $r^*$. Suppose that

$$
S(r^*) = \begin{pmatrix}
15 & 0 & 37 \\
30 & 0 & 8 \\
7 & 0 & 3
\end{pmatrix}.
$$

(28)

Perform the following process on $S(r^*)$ until all the indices 1, 2 and 3 are mapped. First, find the maximum mapping number $s_{13} = 37$. Map the index 1 to 3 and scratch out the first row and the third column from $S(r^*)$. Next find the maximum number $s_{21} = 30$ among the remainder of $S(r^*)$. Then map the index 2 to 1 and scratch out the second row and the first column from $S(r^*)$. After that there exists only one number $s_{32}$ in $S(r^*)$. Map the index 3 to 2 finally. A mapping $M^*$ with $M^*(1) = 3$, $M^*(2) = 1$ and $M^*(3) = 2$ is obtained as a result, which means that the clusters $\Theta_1$, $\Theta_2$ and $\Theta_3$ in the real data $X$ correspond to $\Theta_3^*$, $\Theta_1^*$ and $\Theta_2^*$ in the clustering result $r^*$, respectively.

The clustering mapping method based on the maximum mapping number criterion could be described as follows.

**Cluster Mapping Method by Maximum Mapping Number Criterion**

**Step 0** Set the counter $t = 0$.
**Step 1** Count all the mapping numbers $s_{ik}$ for $i, k \in I$, and obtain the mapping matrix $S$.
**Step 2** Find the maximum number $s_{it_{ik_t}}$ in $S$. Map the the index $i_t$ to $k_t$ and then scratch out the $i_t$th row and the $k_t$th column from $S$.
**Step 3** If $t < c - 1$, then increment $t$ and go to Step 2. Otherwise stop.

As a result of the mapping process above, a mapping $M$ between $\{\Theta_1, \ldots, \Theta_c\}$ and $\{\Theta_1^*, \ldots, \Theta_c^*\}$ is obtained. A point $x_j$ in $X$ is said to be labelled correctly if $M(l_j) = r_j$, where $l_j$ is the label of $x_j$ in the real data of $X$.

**Example 2.** A clustering result of an image with three clusters — mount, river and village — is shown in Figure 2. In order to decide the real cluster of each color in Figure 2(b), some points labelled with different colors, for example, 100 points, are selected randomly according to the proportions, and then identified by going to the actual situation to find out the real clusters. Suppose that 50 points labelled with black in Figure 2(b) correspond to the mount, and 30 points with grey in Figure 2(b) correspond to the river in the real situation. Then the cluster with the black color in Figure 2(b) will be viewed as the real mount, the cluster with grey represents the real river, and the cluster with white is regarded as the real village. The process above is performed according to the maximum mapping number criterion substantially.
Remark 3. The maximum mapping number criterion is used if the mapping number is thought to be prior to the other factors by the users. It is sure that there are many other criteria for deciding the mapping between \( \{\Theta_1, \ldots, \Theta_c\} \) and \( \{\Theta_1^*, \ldots, \Theta_c^*\} \) except the maximum mapping number criterion. For example, if the cluster centers \( a_i^*, i \in I \) are emphasized by the users, a mapping could be given by assigning the cluster center \( \Theta_i^* \) to the nearest cluster \( \Theta_k \) in \( \{\Theta_1, \ldots, \Theta_c\} \). Different criteria could be adopted in reality according to the application of clustering.

In this paper, the maximum mapping number criterion will be employed to obtain the mapping for all the clustering results.

### 4.2. Evaluation of clustering results

There are many evaluation criteria for a clustering result \( \mathbf{r} \), for example, the overall accuracy, and the real cluster number. The evaluation of a clustering result \( \mathbf{r} \) is often made through the error matrix \( E(\mathbf{r}) \).

After the mapping \( M \) of a clustering result \( \mathbf{r} \) is decided, rearrange the columns of the mapping matrix \( S(\mathbf{r}) \) by the order of \( M(1), \ldots, M(c) \), and we get the error matrix \( E(\mathbf{r}) \) of the clustering result \( \mathbf{r} \) with

\[
E(\mathbf{r}) = \begin{pmatrix}
    e_{11} & e_{12} & \cdots & e_{1c} \\
    e_{21} & e_{22} & \cdots & e_{2c} \\
    \vdots & \vdots & \ddots & \vdots \\
    e_{c1} & e_{c2} & \cdots & e_{cc}
\end{pmatrix}.
\]

When valuating a clustering result, the overall accuracy is often used as the most important criterion. The overall accuracy \( \varphi(\mathbf{r}) \) of a clustering result \( \mathbf{r} \) is the proportion of points correctly labelled in \( \mathbf{r} \) among all the points in \( X \), which is
calculated by

$$\varphi(r) = \frac{\sum_{i=1}^{n} e_{ii}}{n},$$  \hspace{1cm} (30)

where $e_{ii}, i \in I$ are the main diagonal elements in the error matrix $E(r)$.

As another important character of a clustering result $r$, the real cluster number $c(r)$ means the number of indices involved in $r$. It is clear that $c(r)$ is just the number of columns with at least one non-zero number in the error matrix $E(r)$.

**Example 3.** The error matrix of the clustering result $r^*$ in Example 2 is

$$E(r^*) = \begin{pmatrix} 37 & 15 & 0 \\ 8 & 30 & 0 \\ 3 & 7 & 0 \end{pmatrix},$$  \hspace{1cm} (31)

the overall accuracy $\varphi(r^*) = (37 + 30 + 0)/100 = 67\%$, and the real cluster number $c(r^*) = 2$.

**Remark 4.** A clustering result $r$ with the real cluster number $c(r) = 1$ is unsatisfactory for sure even though its overall accuracy is high. In order to see that, let us look at Figure 3, in which (a) represents the correct label clustering of Figure 2(a), and (b) is a clustering result with only one cluster. In this result, the only cluster detected will correspond to the mount according to the maximum mapping number criterion. The fact that the proportion of the mount in Figure 2(a) is 72% implies that the overall accuracy of Figure 3(b) is 72%. However, it is obvious that the clustering result in Figure 3(b) is very poor since we cannot obtain any cluster information about the image from this result. This example also reveals that evaluating clustering results only by the overall accuracy is not appropriate in some situations.

**Remark 5.** More clusters do not imply absolutely better clustering results. For example, let us look at other two clustering results in Figure 2(a), one of the results is shown in Figure 4(a) with three clusters and overall accuracy 35%, and the other is in Figure 4(b) with two clusters and overall accuracy 95%. It is obvious that the result in Figure 4(b) is better than that in Figure 4(a).

In many cases, measuring all the results only by the overall accuracy and the real cluster number is not enough. Thus, more evaluation criteria have been proposed in the literature. For example, the user's and producer's accuracies, the kappa value, and the Tau coefficient. Although additive criteria could always lead to more reasonable decision when valuating clustering results, but for our purpose, only the overall accuracy and the real cluster number will be taken into account when we compare all the clustering results of various algorithms in this paper.
4.3. Iris data

The Iris data set has \( n = 150 \) points in a \( p = 4 \)-dimensional space belonging to \( c = 3 \) clusters,\(^{15,16}\) each with 50 points. Among the three clusters, two of the clusters have substantial overlapping.

In order to cluster the Iris data set efficiently with easy performance, two possibilistic clustering algorithms \( \text{PCA}\{f_1\} \) and \( \text{PCA}\{f_2\} \) are presented and compared with PCA93, PCA96 and PCA06, where

\[
    f_1(x) = \frac{1}{1 + 2x^2}, \quad f_2(x) = 2^{-x^2}. \tag{32}
\]

We also run the the fuzzy \( c \)-means clustering algorithm for the Iris data so as to compare it with all the PCAs. For each algorithm, 1000 experiments with random initial centers are implemented to investigate the accuracy distribution of
these clustering algorithms. The small number $\epsilon$ used in the convergence condition $\max_i ||a_i^{(t+1)} - a_i^{(t)}|| < \epsilon$ is set as 0.0001 in all the algorithms. We call it the convergence constant in the following context.

The analysis on accuracies of all the clustering results with various fuzzifier values are given in Table 1, where the columns “Highest”, “Mean” and “Variance” below “Overall Accuracy $\varphi$” provide the highest one, mean and variance of the overall accuracies of 1000 experiments for each algorithm, respectively. In order to show the accuracy distribution of 1000 clustering results, the numbers of experiments with overall accuracies in the regions $[0.8, 1.0]$, $[0.6, 0.8]$ and $[0.0, 0.6]$ are counted and given in the three columns below “Accuracy Distribution”. In addition, the overall accuracy results for $m = 2.0$ are sketched as a histogram shown in Figure 5, in which the height represents the number of experiments with various overall accuracies for each algorithm. The last three columns below “Real CluNum $c$” show the number of clusters obtained among 1000 experiments for $c = 1, 2, 3$, respectively. In both PCA93 and PCA96, the parameters $\eta_i$ were reestimated in each iteration to obtain clustering results with higher overall accuracies, and the parameter $K$ was set to 1.

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Overall Accuracy $\varphi$ (%)</th>
<th>Accuracy Distribution</th>
<th>Real CluNum $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highest</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>$m = 2.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>92.667</td>
<td>80.005</td>
<td>281.67</td>
</tr>
<tr>
<td>PCA96</td>
<td>95.333</td>
<td>77.231</td>
<td>179.36</td>
</tr>
<tr>
<td>PCA06</td>
<td>92.000</td>
<td>79.647</td>
<td>263.24</td>
</tr>
<tr>
<td>PCA{$f_1$}</td>
<td>92.667</td>
<td>80.011</td>
<td>274.04</td>
</tr>
<tr>
<td>PCA{$f_2$}</td>
<td>92.000</td>
<td>79.277</td>
<td>266.47</td>
</tr>
<tr>
<td>FCM</td>
<td>89.333</td>
<td>89.265</td>
<td>1.538</td>
</tr>
<tr>
<td>$m = 2.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>92.000</td>
<td>66.156</td>
<td>4.036</td>
</tr>
<tr>
<td>PCA96</td>
<td>95.333</td>
<td>78.634</td>
<td>162.11</td>
</tr>
<tr>
<td>PCA06</td>
<td>92.000</td>
<td>79.347</td>
<td>270.01</td>
</tr>
<tr>
<td>PCA{$f_1$}</td>
<td>93.333</td>
<td>80.979</td>
<td>271.35</td>
</tr>
<tr>
<td>PCA{$f_2$}</td>
<td>92.000</td>
<td>80.112</td>
<td>251.04</td>
</tr>
<tr>
<td>FCM</td>
<td>90.000</td>
<td>89.837</td>
<td>3.788</td>
</tr>
<tr>
<td>$m = 3.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>66.000</td>
<td>66.000</td>
<td>0.000</td>
</tr>
<tr>
<td>PCA96</td>
<td>95.333</td>
<td>77.007</td>
<td>189.98</td>
</tr>
<tr>
<td>PCA06</td>
<td>92.000</td>
<td>80.054</td>
<td>276.65</td>
</tr>
<tr>
<td>PCA{$f_1$}</td>
<td>94.667</td>
<td>81.763</td>
<td>285.95</td>
</tr>
<tr>
<td>PCA{$f_2$}</td>
<td>92.667</td>
<td>80.495</td>
<td>278.69</td>
</tr>
<tr>
<td>FCM</td>
<td>90.000</td>
<td>89.790</td>
<td>4.861</td>
</tr>
</tbody>
</table>
Like PCA06, PCA\{f_1\} and PCA\{f_2\} are more easily implemented. From Table 1, we can see that the means of the overall accuracies of results of PCA\{f_1\} and PCA\{f_2\} are better than those of PCA93, PCA96 and PCA06.

Comparing FCM with all the PCAs, we can see that though it is possible for the possibilistic clustering algorithms to provide clustering results with higher overall accuracies than FCM (the highest one even achieves 95.333\% by PCA96), however, the clustering performance of all the PCAs including PCA93, PCA96, PCA06, PCA\{f_1\} and PCA\{f_2\} could not ensure a good stable result. It follows from Table 1 that the variances of the overall accuracies of the PCAs are much larger than those of FCM. The least variance of overall accuracies of the PCAs is 162.11, and the maximum variance of FCM is 4.861 (the overall accuracies of PCA93 with \(m = 2.5\) and 3.0 nearly keep constant since only two clusters are detected). Figure 5 implies that at most 64\% of experiments of the PCAs could provide the clustering results with overall accuracies larger than 80\%, whereas the FCM algorithm presented results with overall accuracy larger than 89\% in more than 98\% of experiments.

In addition, when taking account of real cluster numbers of results, it seems that the possibilistic clustering algorithms are not very accurate at distinguishing the two clusters with overlapping in the Iris data set. For all the PCAs, more than 20\% of experiments could not detect the three clusters completely.

### 4.4. Glass data set

The Glass data set from Blake and Merz has \(n = 214\) points in a \(p = 9\)-dimensional space.\(^{17}\) It consists of \(c = 6\) clusters, which have 70, 76, 17, 13, 9 and 29 points in each cluster, respectively.
In this example, two possibilistic clustering algorithms $\text{PCA}_3$ and $\text{PCA}_4$ are used for clustering the Glass data set, where

$$f_3(x) = \frac{1}{1 + 5x^3}, \quad f_4(x) = 2^{-8x^2}. \tag{33}$$

The clustering performance of $\text{PCA}_3, \text{PCA}_6, \text{PCA}_3, \text{PCA}_4$ and FCM through 1000 experiments is listed in Table 2, where the convergence constant $\epsilon$ was set to 0.00001. The analysis on the overall accuracy and the real cluster number is summarized in Table 2, and the overall accuracy distribution of 1000 experiments for each clustering algorithm with $m = 2.0$ is shown in Figure 6 by a histogram. In $\text{PCA}_3$ and $\text{PCA}_6$, the parameter $K$ was set to 1, and the parameters $\eta_i$ were constant in all the iterations.

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Overall Accuracy $\varphi$ (%)</th>
<th>Accuracy Distribution</th>
<th>Real CluNum $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highest</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>$m = 2.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>48.131</td>
<td>42.714</td>
<td>9.005</td>
</tr>
<tr>
<td>$\text{PCA}_6$</td>
<td>47.664</td>
<td>40.161</td>
<td>11.300</td>
</tr>
<tr>
<td>$\text{PCA}_0$</td>
<td>53.738</td>
<td>40.679</td>
<td>18.839</td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>57.009</td>
<td>48.592</td>
<td>19.028</td>
</tr>
<tr>
<td>$\text{PCA}_4$</td>
<td>56.075</td>
<td>46.200</td>
<td>15.510</td>
</tr>
<tr>
<td>FCM</td>
<td>55.140</td>
<td>49.120</td>
<td>2.655</td>
</tr>
<tr>
<td>$m = 2.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>47.196</td>
<td>39.989</td>
<td>4.894</td>
</tr>
<tr>
<td>$\text{PCA}_6$</td>
<td>47.196</td>
<td>37.489</td>
<td>9.190</td>
</tr>
<tr>
<td>$\text{PCA}_0$</td>
<td>62.617</td>
<td>45.621</td>
<td>16.707</td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>57.477</td>
<td>49.708</td>
<td>12.854</td>
</tr>
<tr>
<td>$\text{PCA}_4$</td>
<td>62.150</td>
<td>46.150</td>
<td>23.708</td>
</tr>
<tr>
<td>FCM</td>
<td>55.140</td>
<td>49.778</td>
<td>3.140</td>
</tr>
<tr>
<td>$m = 3.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>46.262</td>
<td>38.927</td>
<td>0.966</td>
</tr>
<tr>
<td>$\text{PCA}_6$</td>
<td>45.794</td>
<td>35.752</td>
<td>6.284</td>
</tr>
<tr>
<td>$\text{PCA}_0$</td>
<td>56.075</td>
<td>48.400</td>
<td>18.344</td>
</tr>
<tr>
<td>$\text{PCA}_3$</td>
<td>55.607</td>
<td>48.819</td>
<td>11.049</td>
</tr>
<tr>
<td>$\text{PCA}_4$</td>
<td>61.215</td>
<td>45.076</td>
<td>21.771</td>
</tr>
<tr>
<td>FCM</td>
<td>55.140</td>
<td>48.841</td>
<td>2.809</td>
</tr>
</tbody>
</table>

It follows from Table 2 that $\text{PCA}_3$ and $\text{PCA}_4$ have the same efficiency as $\text{PCA}_3, \text{PCA}_6$ and $\text{PCA}_0$ for clustering the Glass data set. The means of overall accuracy of results of $\text{PCA}_3$ keep higher than 48.5% for $m = 2.0, 2.5$ and 3.0, which are much better than the other PCAs.
From another point of view, compared with FCM, all the PCAs in this example could not implement the clustering on the Glass data set with very good stable results, though sometimes they may return better results than FCM (the highest accuracy achieves 62.150% by PCA\( f_4 \), while the highest one by FCM is 55.140%). The clustering results of FCM have the overall accuracies larger than 48% in nearly 97% of experiments, and about 91% of experiments can detect all the six clusters.

4.5. Vowel data set

The Vowel data set from Blake and Merz has \( n = 990 \) points in a \( p = 10 \)-dimensional space.\(^{17}\) It consists of \( c = 11 \) clusters, each with 90 points.

Similarly, we provide two possibilistic clustering algorithms PCA\( f_5 \) and PCA\( f_6 \) for clustering the Vowel data set, and compare with PCA93, PCA96, PCA06 and FCM, where

\[
f_5(x) = \frac{1}{1 + 20x^2}, \quad f_6(x) = \exp\{-x^2\}. \tag{34}
\]

1000 experiments were run for each algorithm with convergence constant \( \epsilon = 0.000001 \) and fuzzifier \( m = 2.0, 2.5 \) and 3.0, respectively, and the summary on the overall accuracy and real cluster number analysis of all the experiments are given in Table 3. The overall accuracy distribution of 1000 experiments for each clustering algorithm with \( m = 2.0 \) is shown in Figure 7 by a histogram. The parameter \( K \) in both PCA93 and PCA96 was set to 1. In order to obtain the better clustering results, the parameters \( \eta_i \) were unchanged in all the iterations in both PCA93 and PCA96.
Table 3. A comparison of 1000 experiments for the Vowel data.

<table>
<thead>
<tr>
<th>Clustering Algorithm</th>
<th>Overall Accuracy $\varphi$ (%)</th>
<th>Accuracy Distribution</th>
<th>Real CluNum $c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highest</td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>$m = 2.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>32.525</td>
<td>23.428</td>
<td>8.172</td>
</tr>
<tr>
<td>PCA96</td>
<td>32.727</td>
<td>23.488</td>
<td>5.848</td>
</tr>
<tr>
<td>PCA06</td>
<td>39.899</td>
<td>30.939</td>
<td>7.572</td>
</tr>
<tr>
<td>${f_5}$</td>
<td>37.879</td>
<td>28.781</td>
<td>9.418</td>
</tr>
<tr>
<td>${f_6}$</td>
<td>39.192</td>
<td>31.858</td>
<td>7.211</td>
</tr>
<tr>
<td>FCM</td>
<td>32.020</td>
<td>28.001</td>
<td>1.151</td>
</tr>
<tr>
<td>$m = 2.5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>36.061</td>
<td>25.152</td>
<td>9.223</td>
</tr>
<tr>
<td>PCA96</td>
<td>31.515</td>
<td>24.025</td>
<td>5.575</td>
</tr>
<tr>
<td>PCA06</td>
<td>39.697</td>
<td>30.473</td>
<td>6.898</td>
</tr>
<tr>
<td>${f_5}$</td>
<td>38.990</td>
<td>30.473</td>
<td>7.058</td>
</tr>
<tr>
<td>${f_6}$</td>
<td>40.404</td>
<td>31.020</td>
<td>7.097</td>
</tr>
<tr>
<td>FCM</td>
<td>32.727</td>
<td>23.071</td>
<td>6.440</td>
</tr>
<tr>
<td>$m = 3.0$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA93</td>
<td>35.556</td>
<td>26.103</td>
<td>9.652</td>
</tr>
<tr>
<td>PCA96</td>
<td>33.636</td>
<td>24.411</td>
<td>5.539</td>
</tr>
<tr>
<td>PCA06</td>
<td>40.101</td>
<td>30.476</td>
<td>6.698</td>
</tr>
<tr>
<td>${f_5}$</td>
<td>39.899</td>
<td>30.536</td>
<td>6.669</td>
</tr>
<tr>
<td>${f_6}$</td>
<td>40.404</td>
<td>30.776</td>
<td>6.798</td>
</tr>
<tr>
<td>FCM</td>
<td>35.051</td>
<td>24.050</td>
<td>7.120</td>
</tr>
</tbody>
</table>

It follows from Table 3 that PCA$\{f_5\}$ provided clustering results with the average accuracy 31.858% when $m = 2.0$, which is the best result of all the PCAs and FCM. It can be also seen that PCA06, PCA$\{f_5\}$ and PCA$\{f_6\}$ detected 11 clusters in more than 91% of experiments when $m = 2.5$ and 3.0, but FCM failed to do that. It seems that some PCAs could provide better results than FCM for the Vowel data set when only taking account of the overall accuracy and real cluster number, although all the results are unsatisfied.

4.6. Summary

Finally, we summarize the result analysis of all the numerical experiments on the possibilistic clustering algorithms in Tables 1–3 by emphasizing the following points:

(a) Compared with PCA06 and the other PCAs like PCA$\{f\}$, PCA93 and PCA96 are not very easy to perform in reality since it is difficult to decide whether to reestimate the parameters $\eta_i$ or not, though they could provide clustering results with high overall accuracy in some situations by using appropriate parameter $K$ or $\alpha$ in (12).
(b) PCA06 is a good clustering algorithm among the existing PCAs because of its easy control for diverse real applications.

(c) Except PCA93, PCA96 and PCA06, there are many other possibilistic clustering algorithms like PCA\{f\} with easy and effective clustering performance, where f satisfies the constraints in (17).

(d) It is possible, but not very stable, for the possibilistic clustering algorithms to provide clustering results with high overall accuracies, especially when clustering the Iris and Glass data sets.

**Remark 6.** The conclusions drawn above are based on the clustering results of the three real data sets experiments. In order to extend these conclusions, theoretic analysis or experiments on more data sets are necessary.

5. Conclusion

Fuzzy clustering is an approach using the fuzzy set theory as a tool for data grouping. In the literature, many fuzzy clustering algorithms based on objective function have been proposed including FCM and some PCAs.

It is well-known that the memberships resulting from FCM do not always correspond to the intuitive concept of degree of belongingness or compatibility. Thus, though the FCM algorithm does give meaningful results in applications as probabilities or degree of sharing,\(^{20}\) it is not suitable for applications in which the memberships are supposed to represent typicality or compatibility.\(^{21,22}\) On the contrary, the memberships generated by all the PCAs provide a good explanation of degrees
of belongingness for the data. Thus, they are efficient for the situations in the presence of noise in the data, which has been discussed by Krishnapuram and Keller in detail.\textsuperscript{2}

In this paper, we have contributed to the research area of fuzzy clustering in the following aspects:

(i) A new approach of fuzzy clustering was suggested by computing the memberships directly using the fuzzy set theory according to the data information, and updating the cluster centers via a performance index.

(ii) A generalized approach to PCAs was presented, which leads to a new infinite family of PCAs involving the existing PCAs as members.

(iii) Some real data experiments on the basis of specific possibilistic clustering algorithms in the new family were given. The results compared with the existing PCAs showed the new proposed algorithms are efficient for clustering these real data sets with easy performance.

(iv) The comparison of the PCAs and FCM from the accuracy and real cluster number was accomplished based on some numerical experiments. The results revealed that the PCAs could not always provide very stable results when clustering the Iris and Glass data sets.

Stability of algorithms is necessary for real applications. From the results of numerical experiments in Section 4, we see that the PCAs are not very stable when clustering the Iris and Glass data sets compared with FCM. However, the clustering results for the Vowel data set are inverse. It is really challenging to explore whether the PCAs are unstable through theoretic analysis or more real data experiments. As another important issue, further discussion on the membership evaluation equation must be done so as to obtain more efficient clustering results in the future research on the possibilistic clustering algorithms, in which the concept of entropy of fuzzy variables may be used.\textsuperscript{23}

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 60425309). We are grateful to our students, S. Arasteh, S. Saatchi, M. Zhang, M. Xiang and D.-K. Chen for their suggestions and excellent assistance in numerical experiments.

References